# Discretization of Self-Exciting Peaks Over Threshold Models*

Daisuke Kurisu†

**Abstract**

In this paper, a framework on a discrete observation of (marked) point processes under the high-frequency observation is developed. Based on this framework, we first clarify the relation between random coefficient integer-valued autoregressive process with infinite order (RCINAR($\infty$)) and i.i.d.-marked self-exciting process, known as marked Hawkes process. For this purpose, we show that the point process constructed of the sum of a RCINAR($\infty$) converge weakly to a marked Hawkes process. This limit theorem establish that RCINAR($\infty$) processes can be seen as a discretely observed marked Hawkes processes when the observation frequency increases and thus build a bridge between discrete-time series analysis and the analysis of continuous-time stochastic process and give a new perspective in the point process approach in extreme value theory. Second, we give a necessary and sufficient condition of the stationarity of RCINAR($\infty$) process and give its random coefficient autoregressive (RCAR) representation. Finally, as an application of our results, we establish a rigorous theoretical justification of self-exciting peaks over threshold (SEPOT) model, which is a well-known as a (marked) Hawkes process model for the empirical analysis of extremal events in financial econometrics and of which, however, the theoretical validity have rarely discussed. Simulation results of the asymptotic properties of RCINAR($\infty$) shows some interesting implications for statistical applications.

**Key Words**

Peaks over threshold, Extreme value theory, Marked Hawkes process, High-frequency data, RCINAR($\infty$) process.

## 1 Introduction

Peaks over threshold (POT) method have been used in a large number of scientific fields in the last decades, and the methodology is related to the theory of extreme value

*This version: December 21, 2016.

†Graduate student, Graduate School of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: dkurisu.mathstat@gmail.com

analysis and point process. In the literature of extreme value theory, generalized Pareto distributions (GPD) have been investigated since the early works of Pickands (1975) and Balkema and de Haan (1974). According to their results, GPD is a natural distributional model of the data that exceeds a particular designed high level. Resnick (1987) introduced point process approach for extreme value analysis and developed the modeling of threshold exceedance based on the point process theory. The classical POT models assume that those extremal events occur according to a time homogeneous Poisson process. but this assumption of time homogeneity is not satisfied in some applications

Chavez-Demoulin, Davison and McNeil (2005) generalized this idea to the case that the rate of exceedance follows a self-exciting point process (Hawkes process), called self-exciting peaks over threshold (SEPOT) model. We refer to Chavez-Demoulin and McGill (2012), Grothe, Korniichuk and Manner (2014), and Chavez-Demoulin, Embrechts and Sardy (2014) for recent contributions on this topic. Hawkes process is a class of point process developed in Hawkes (1971) and have been investigated in a wide range of fields such as seismology, neural science, and finance. The notable characteristic of Hawkes process is the construction of intensity process which depends on the past of process itself. Liniger (2009) gave a rigorous construction of Hawkes process, and mathematical properties are discussed in the paper. SEPOT model is applied to a data that assumed to be a realization of a (marked) point process in most cases of data analysis in financial econometrics.

However, such an assumption may not be realistic if we regard the observed data are discrete observations of a (marked) point process in the case when the observation frequency increases. In the literature of high-frequency data analysis, it is usually assumed that the observed data is a discrete observation of some underlying continuous-time stochastic process: An underlying stochastic process $X = (X_t)_{t\in\mathbb{R}}$, which is a semimartingale, is sampled at discrete times $k\Delta$, $k \in \mathbb{Z}$ and we are available the values $\{X_{k\Delta}, k \in \mathbb{Z}\}$. The increments $\{\Delta_k X = X_{k\Delta} - X_{(k-1)\Delta} : k \in \mathbb{Z}\}$ are the discretized process of an underlying stochastic process, and point process modeling is performed to these increments in the most of empirical analysis of financial time series. In the case of high-frequency observation, that is, the mesh $\Delta$ goes to 0, it may be reasonable to consider the values $\{X_{k\Delta}, k \in \mathbb{Z}\}$ as the discretely observed point process since a marked point process is a class of semimartingale (see Aït-Sahalia and Jacod (2014) for the theoretical framework of high-frequency data analysis).

To fill this discrepancy between theory and empirical applications, we develop a new framework on the discrete observation of (marked) point processes and investigate the discretization of SEPOT models. We then establish a theoretical justification of SEPOT methods whihc are often used in empirical analyses. The key idea of our paper is the discretization of (marked) Hawkes process under the high-frequency observation. Kirchner (2016a) studied the discrete-time approximation of Hawkes process and provides a

limit theorem that integer-valued auto-regressive process with infinite order (INAR($\infty$)) approximately can be seen as a increments of Hawkes processes. This contributes to link the ideas which have been discussed in the two different contexts, the classical statistical time series analysis and the probabilistic analysis of continuous-time stochastic process. Our results are generalization of the results in Kirchner (2016a) to marked Hawkes process, and give a new perspective in the point process approach in extreme value theory.

INAR($p$) process is first defined McKenzie (1985) in the context of classical discrete time series analysis for the models of count data. The theoretical development is made in Al-Osh and Alzaid (1987) and Du and Li (1991). Zheng, Basawa and Datta (2007) introduced INAR(1) to the random coefficient (RCINAR(1)) case and extended in Zhang, Wang and Zhu (2011a) and Zhang, Wang and Zhu (2011b). Boshnakov (2011) investigated the stationarity of RCINAR($p$). We refer to McCabe, Martin and Harris (2011) and McKenzie (2003) for a general review of integer-value time series models and their applications. We give some properties of RCINAR($\infty$) and RCINAR($p$), and give a limit theorem on the weak convergence of RCINAR($\infty$) to the corresponding marked Hawkes process.

We also checked finite sample properties of our results by numerical experiments. In our simulation, we considered two cases: self-exciting and self-damping. The latter case means the situation that the occurrence of an event decreases the probability of occurrence of the event in the future, and such a situation can seen in the field of seismology and economics. Simulation results show that the approximation of RCINAR($\infty$) by RCINAR($p$) is sensitive to the choice of both $p$ and $\Delta$.

The construction of this paper is as follows: In Section 2 we introduce a general framework on a discrete observation of (marked) point process. In Section 3, we explain a basic result on marked point process which related to extreme value theory, and give some examples of SEPOT models. In Section 4, we give some properties of RCINAR($\infty$) and show that RCINAR($\infty$) can be seen as a local approximation of a marked Hawkes intensity process. We confirm finite sample properties of our results through simulations in Section 5. Some directions of extension and application of our results are discussed in Section 6. We conclude this paper in Section 7. Proofs are gathered in Appendix A.

## 2 Discrete Observation of Point Process

In this section we introduce a new framework on a discrete observation of point process models. Most of papers on an empirical analysis of financial time series in which (marked) point process models are used treat the observed data as a realization of a latent (marked) point process. However, such an assumption on the data is not realistic if we stand in a position that the observations are discrete samples of an underlying

point process. For example, in financial econometrics, increasing interests have been paid on high-frequency data analysis of asset prices which is observed typically in every one seconds. In high-frequency financial data analysis, it is usually assumed that the observed data are discrete observations of a semimartingale, or a continuous time stochastic process for the mathematical treatment of observations. Therefore, there is a void in high-frequency data analysis between mathematical assumptions based on a semimartingale theory and empirical applications based on point process models. For this reason, we attempt to develop a mathematical framework that links these two different points of view in this paper. First we introduce a general framework on the discrete observation of point processes, then we focus on the discrete observation of marked self-exciting process, known as marked Hawkes process. Let $N$ be a one dimensional point process and its latent (generally unobserved) event times in the observation interval $[0, T]$, $0 \leq t_1 < t_2 < \cdots < t_{N_T} \leq T$. If the observation distance is $\Delta$, we define the number of jumps in each observation intervals

$$J_k = \#\{i : t_i \in ((k-1)\Delta, k\Delta]\}.$$

We consider a situation that an underlying point process is sampled at high-frequency, that is, $\Delta \to 0$. In this case, for sufficiently small $\Delta$, we have $J_k \in \{0, 1\}$, $k = 1, \ldots, [T/\Delta]$. We also consider auxiliary stochastic processes

$$N_t^{\Delta} = N_{(k-1)\Delta}, \quad (k-1)\Delta \leq t < k\Delta.$$

We have the following relation for each observation interval:

$$\sup_{(k-1)\Delta \leq t < k\Delta} |N_t^{\Delta} - N_t| \leq J_k.$$

If we take the upper bound of the observation distance $\Delta_0$ sufficiently small, we have

$$\sup_{0 < \Delta < \Delta_0} \sup_{0 \leq t < T} |N_t^{\Delta} - N_t| \leq 1$$

in general. For the fitting of point process models to observed data at times $0 \leq \Delta < 2\Delta < \ldots$, we set a threshold value $u_0$ and the occurrence of an event is recognized if the observed value at time $k\Delta$ exceeds the threshold $u_0$. Such events are assumed to be a realization of an under lying point process in empirical financial time series analysis. To explain this assumption mathematically, we introduce the following stochastic process:

$$\widetilde{N}_t^{\Delta} = \sum_{1 \leq l \leq k-1} \frac{N(((l-1)\Delta, l\Delta])}{J_l}, \quad (k-1)\Delta \leq t \leq k\Delta.$$

The stochastic process $\widetilde{N}^{\Delta}$ varies at most 1 at each observation time $k\Delta$. In most of empirical studies, $\widetilde{N}^{\Delta}$ is regarded as a realization of point process $N$, and if we can

only available the values at $k\Delta$, it is considered that no events occurred in the interval $((k-1)\Delta, k\Delta]$ in the case when the value at time $k\Delta$ does not exceeds the predetermined threshold $u_0$. In this case, the events occurred in $((k-1)\Delta, k\Delta)$ is ignored. However, if we can also available the maximum or minimum values in $((k-1)\Delta, k\Delta]$, such ignorances do not happen, that is, we can set

$$\begin{aligned} J_k > 0 &\Leftrightarrow N \text{ have at most one event in } ((k-1)\Delta, k\Delta] \\ &\Leftrightarrow \widetilde{N}^{\Delta}_{k\Delta} - \widetilde{N}^{\Delta}_{(k-1)\Delta} = N^{\Delta}_{k\Delta} - N^{\Delta}_{(k-1)\Delta} = 1. \end{aligned}$$

Intuitively, we can approximate $N$ by $\widetilde{N}^{\Delta}$ when $\Delta \to 0$. We establish the theoretical validity of this approximation in Section 4.

# 3 Preliminaries

In this section, we see the relation between point process and extreme value theory for the purpose of the discretization of SEPOT models discussed in Section 4. In Section 3 and Section 4, we consider the weak convergence of point processes. Hence we first explain the concept.

Let $E$ be a complete separable metric space, $M_p(E)$ be the space of all point measures defined on $E$, and $C^+_K(E)$ be the family of all non-negative continuous functions defined on $E$ with compact support. We can construct a topology on $M_p(E)$ based on the concept of vague convergence: For $\mu_n, \mu \in M_p(E)$, we say $\mu_n$ converge vaguely to $\mu$ if $\mu_n(f) = \int_E f d\mu_n \to \mu(f) = \int_E f d\mu$ for any $f \in C^+_K(E)$. It is known that this topology on $M_p(E)$ is metrizable as a complete separable metric space (Proposition 3.17 in Resnick (1987)). Let $\mathcal{M}_p(E)$ be the Borel $\sigma$-field on $M_p(E)$. A Point process is defined as a random variable from some stochastic space $(\Omega, \mathcal{A}, P)$ to the measurable space $(M_p(E), \mathcal{M}_p(E))$. Therefore, we can use the argument of weak convergence of random variables on a metric space developed in Billingsley (1968). It is also known that for (marked) point processes, the weak convergence of a point process is equivalent to the weak convergence of finite dimensional distributions are equivalent (Theorem 11.1.IV in Daley and Vere-Jones (2003)). In the following sections, we use the notation $\overset{w}{\to}$ as weak convergence.

## 3.1 Peaks Over Threshold Method

Let $(X_i)$ be an i.i.d. sequence of random variables with distribution function $F$ and $D(G_\xi)$ be the domain of attraction of a generalized extreme value distribution with parameter $\xi$. Then $F \in D(G_\xi)$ means that there exist a centering sequence $\{a_n\}$, $a_n \in \mathbb{R}$ and a scaling sequence $\{b_n\}$, $b_n > 0$ of an extreme value distribution $G_\xi$ such that $mP(b_m^{-1}(X_1 - a_m) \geq u) = m(1 - F(b_m u + a_m)) \to (1 + \xi u)^{-1/\xi}$ as $m \to \infty$ for any

$u > 0$. The random variable $Z$ with distribution function

$$F(x) = 1 - \left(1 + \xi \frac{x^+}{\sigma}\right)^{-1/\xi}, \sigma > 0, \xi \in \mathbb{R}, \quad a^+ = \max(a, 0),$$

is called a generalized Pareto distributed random variable and we write this as $Z \sim GPD(\xi, \sigma)$ (see de Haan and Ferreira (2006) for details of extreme value distributions). GPD is known as a natural distributional model for the data that exceeds a particularly designed high level (McNeil, Frey and Embrechts (2005)). Proposition 1 is the well known result on the peaks over threshold method.

**Proposition 1** (Theorem 6.3 in Resnick (2007))**.** *Let $(X_k)_{1 \le k \le [\Delta_n^{-1}]}$ be an i.i.d. sequence with distribution $F \in D(G_\xi)$. Consider the following marked point process*

$$N_E^{\Delta_n}((a, b]) = \# \left\{ k : \left( k\Delta_n, \frac{X_k - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \right) \in (a, b] \times E, \right\}, a < b, E = [u, \infty), u \ge 0,$$

*where $\{a_n\}$ and $\{b_n\}$ are centering and scaling sequences of an extreme value distribution $G_\xi$. Then $N_E^{\Delta_n} \xrightarrow{w} N$ in $M_p([0, \infty) \times E)$ as $\Delta_n \to 0$, where $N$ is a time homogeneous Poisson process with intensity*

$$\lambda = (1 + \xi u)^{-1/\xi}.$$

**Remark 1.** The point process $N_E^{\Delta_n}$ in Propositon 1 can be rewritten in the following form considering the Poisson law of small numbers:

$$N_E^{\Delta_n}((a, b]) = \sum_{k : k\Delta_n \in (a, b]} \widetilde{\epsilon}_n(k) \approx \sum_{k : k\Delta_n \in (a, b]} \epsilon_n(k), \quad E = [u, \infty), u \ge 0,$$

$$\widetilde{\epsilon}_n(k) \overset{i.i.d.}{\sim} Bin\left(1, P\left(\frac{X_k - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \ge u\right)\right),$$

$$\epsilon_n(k) \overset{i.i.d.}{\sim} Poi\left(P\left(\frac{X_k - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \ge u\right)\right),$$

where $Bin(1, \alpha)$ and $Poi(\lambda)$ denote Bernoulli and Poisson random variables with parameters $\alpha$ and $\lambda$ respectively. Therefore, we can see the data $\{k\Delta_n, (X_k - a_{[\Delta_n^{-1}]})/b_{[\Delta_n^{-1}]}\}_{1 \le k \le [\Delta_n^{-1}]}$ as a discrete observation of the Poisson process with intensity $\lambda' = 1$, and the probability that regularized random variables $\{(X_k - a_{[\Delta_n^{-1}]})/b_{[\Delta_n^{-1}]}\}_{1 \le k \le [\Delta_n^{-1}]}$ exceed the threshold $u$ can be approximated by $(1 + \xi u)^{-1/\xi}$, which equals to the probability $P(Z \ge u), Z \sim GPD(\xi, 1)$. Therefore, it is possible to reinterpret the basic result in extreme value theory from the standpoint of high-frequency observation.

Proposition 1 assumes that the occurrence of events are regularly spaced, that is, the times of exceedance of a threshold are equidistant. However, this assumption is often violated by financial time series for instance. Self-exciting POT models overcome this problem, and allows the irregular event times and then enables us to model the clustering of events.

### 3.2 Self-Exciting Peaks Over Threshold Model

The SEPOT model is proposed in Chavez-Demoulin, Davison and McNeil (2005) and since then it has been used for the modeling of extremal events in financial econometrics. This model is defined by a (marked) Hawkes process which is known as self-exciting process, is proposed in Hawkes (1971) and studied in many scientific fields such as biology, neural science, seismology and financial econometrics. It is well known that the law of a $\mathcal{H}_t$-adapted point process $N$ is uniquely determined by its intensity process defined by

$$\begin{aligned}\lambda(t|\mathcal{H}_t) &= \lim_{\Delta\to 0} E\left[\frac{N(t+\Delta)-N(t)}{\Delta}|\mathcal{H}_t\right]\\ &= \lim_{\Delta\to 0}\frac{P(N(t+\Delta)-N(t)>0|\mathcal{H}_t)}{\Delta}.\end{aligned}$$

More precisely, a intensity $\lambda(t|\mathcal{H}_t)$ is determined by a point process with the following properties:

$$\begin{aligned}&P(N(t+\Delta)-N(t)=1|\mathcal{H}_t)=\lambda(t|\mathcal{H}_t)\Delta+o_P(\Delta),\\ &P(N(t+\Delta)-N(t)=0|\mathcal{H}_t)=1-\lambda(t|\mathcal{H}_t)\Delta+o_P(\Delta),\\ &P(N(t+\Delta)-N(t)>1|\mathcal{H}_t)=o_P(\Delta).\end{aligned}$$

This properties implies that the point process $N$ is *simple*, which means that almost all sample path of $N$ have no *co-jumps* (i.e. $N(\{x\}) = \{0,1\}$ a.s. $\forall x \in E$). The intensity based definition of Hawkes process is given as follows:

**Definition 1** (Hawkes intensity process with no marks)**.** *Hawkes process $N$ is a point process with the following intensity:*

$$\lambda(t|\mathcal{H}_t^N) = \eta + \int_{-\infty}^{t} h(t-s)N(ds)$$

*where $\eta > 0$, $h : \mathbb{R} \to \mathbb{R}_+ = [0,\infty)$ is a piecewise continuous function called decay function and $\mathcal{H}_t^N$ is a filtration which contains the canonical filtration $\mathcal{G}_t^N = \sigma(N_s, : s < t)$.*

**Remark 2.** Note that time homogeneous Poisson processes can be seen as a special version of Hawkes processes with no self-excitation, that is, the case when $h \equiv 0$ in Definition 1.

If we have additional information on a point process, the size of exceedance of a threshold for example, we can incorporate it as marks in its intensity process. The intensity process of marked Hawkes process is defined as follows:

**Definition 2** (Marked Hawkes intensity process)**.** *Let $E$ be a complete separable metric space. A marked Hawkes process $N$ with $E$-valued mark is a marked point process with the following intensity:*

$$\begin{aligned}\lambda_E(t|\mathcal{H}_t^N) &= \left(\eta + \int_{-\infty}^{t}\int_E g(t-s,z)N(ds\times dz)\right)P(\widetilde{Z}_t \in E|\mathcal{H}_t^N),\\ &= \left(\eta + \int_{-\infty}^{t} g(t-s,\widetilde{Z}_s)N_g(ds)\right)P(\widetilde{Z}_t \in E|\mathcal{H}_t^N),\end{aligned}$$

*where $N_g(ds) = N(ds \times E)$, $\mathcal{H}_t^N$ is a filtration which contains the canonical filtration $\mathcal{G}_t^N = \sigma(N_s : s < t)$, $\eta > 0$, $g : \mathbb{R}^2 \to \mathbb{R}_+$ is a piecewise continuous function, $(\widetilde{Z}_s)_{s\in\mathbb{R}}$ is an $\mathcal{H}_t^N$ adapted $E$-valued stochastic process associated to an event at time $s$ called the mark of this point process, and $P(\widetilde{Z}_t \in \cdot|\mathcal{H}_t^N)$ is a $\mathcal{H}_t^N$- measurable conditional distribution of $\widetilde{Z}_t$.*

In the point process theory, the point process $N_g$ is called *ground process* (Daley and Vere-Jones (2003)). If the function $g : \mathbb{R}^2 \to \mathbb{R}_+$ can be written in a product form $g(t,z) = h(t)c(z)$, then $h$ is also called decay function and $c : \mathbb{R} \to \mathbb{R}_+ = [0,\infty)$ is called impact function (see Daley and Vere-Jones (2003) for the detailed definition of general marked point process and Liniger (2009) for general mathematical construction of Hawkes processes including marked Hawkes process).

In the application of SEPOT model, we set a threshold value $u_0$ and counts events that exceed the threshold, for example, the event that an asset price fall below some level of negative return. The rate of exceedance also assumed to be driven by Hawkes intensity process with i.i.d. mark defined as follows:

$$\begin{aligned}\lambda_E(t|\mathcal{H}_t^N) &= \left(\eta + \int_{-\infty}^{t}\int_E g(t-s,z)N(ds\times dz)\right)P(Z_1 \in E),\\ &= \left(\eta + \int_{-\infty}^{t} g(t-s,\widetilde{Z}_s)N_g(ds)\right)(1-F(u)), \qquad (1)\end{aligned}$$

where $E = [u,\infty)$, $u \geq 0$, $\widetilde{Z}_s = (Z_s - u)^+ = \max(Z_s - u, 0)$ and $(Z_s)$ is an i.i.d. sequence of real valued random variables with distribution function $F$ associated to event times of $N$. We give some examples of SEPOT models driven by marked Hawkes process.

**Example 1** (Models with exponential decay and generalized linear impact function)**.**

$$\lambda_E(t|\mathcal{H}_t^N) = \left(\eta_0 + \eta_1\int_{-\infty}^{t} e^{-\gamma(t-s)}\left[(\widetilde{Z}_s)^{\delta}\wedge L\right]N_g(ds)\right)(1+\xi u)^{-1/\xi},$$

where $a \wedge b = \min(a,b)$, $\widetilde{Z}_s = (Z_s - u)^+$, $Z_s \overset{i.i.d.}{\sim} \mathrm{GPD}(\xi, 1)$, $\eta_0, \eta_1, \gamma > 0$, $L > 1$ and $0 \leq \delta \leq 1$ are constants. This model includes linear impact function ($\delta = 1$) as a special case. $\mathrm{GPD}(\xi,\sigma)$ is a generalized Pareto distribution. This type of models are considered

in Kunitomo, Ehara and Kurisu (2016). They apply their models for the Granger-non-causality test of international financial market indices. It is also posssible to consider polynomial decay function $h(t) = (\gamma + t)^{-(p+1)} 1_{\{t \geq 0\}}$, $\gamma, p > 0$.

**Example 2** (Models with exponential decay and nonlinear impact function)**.**

$$\lambda_E(t|\mathcal{H}_t^N) = \left(\eta_0 + \eta_1 \int_{-\infty}^{t} e^{-\gamma(t-s)}[(1 + G^{\leftarrow}(F(\widetilde{Z}_s))) \wedge L] N_g(ds)\right)(1 + \xi u)^{-1/\xi},$$

where $\widetilde{Z}_s = (Z_s - u)^+$, $Z_s \overset{i.i.d.}{\sim} \mathrm{GPD}(\xi, 1)$, $\eta_0, \eta_1$, $\gamma$ and $L > 1$ are positive constants, $F$ is a cumulative distribution function of $\mathrm{GPD}(u, \xi, 1)$, and $G^{\leftarrow}(\cdot)$ is the inverse of a distribution function $G$ of some continuous positive random variable with finite mean $\delta$. This type of model is a special case of that considered in Grothe, Korniichuk and Manner (2014). They apply their model for the prediction of probabilities of future jumps of asset prices.

# 4 Main Results

To our knowledge, in contrast to the applicability of the SEPOT model, the theoretical justification of the model have not been established. In this section, we establish the validity of SEPOT models in the aspect of discrete observation of point process. For the description our results, we first introduce random coefficient integer-valued autoregressive process with infinite order (RCINAR($\infty$)).

## 4.1 RCINAR($\infty$) Process

RCINAR(1) process is introduced in Zheng, Basawa and Datta (2007) and generalized to RCINAR($p$) process in Zhang, Wang and Zhu (2011a,b). In this section we introduce the random coefficient integer-valued auto regressive process with infinite order (RCINAR($\infty$)) and describe some basic properties of this model. The definition of RCINAR($\infty$) is defined as follows:

**Definition 3** (RCINAR($\infty$))**.** *Let $(\epsilon_n)$ and $(Z_n)$ be an i.i.d. sequence of random variables with Poisson distribution $Poi(\alpha_0)$, $\alpha_0 > 0$ and distribution $F$, and $\alpha_k : \mathbb{R} \to \mathbb{R}$, $k = 1, 2, \ldots$. A random coefficient integer-valued autoregressive process with infinite order (RCINAR($\infty$)) is a stochastic process given by*

$$\begin{aligned} \epsilon_n &= X_n - \sum_{k=1}^{\infty} \alpha_k(Z_{n-k}) \circ X_{n-k} \\ &= X_n - \sum_{k=1}^{\infty} \sum_{l=1}^{X_{n-k}} \widetilde{\xi}_l^{n,k}, \quad n \in \mathbb{Z}. \end{aligned} \tag{2}$$

*where* $\widetilde{\xi}_l^{n,k} \sim Poi(\alpha_k(Z_{n-k}))$ *and independent in* $l, n$ *and* $k$. *For* $\alpha < 0$, *we interpret* $\alpha \circ X = -(-\alpha) \circ X$.

The next proposition gives a necessary and sufficient condition for the stationarity of RCINAR($\infty$) process.

**Proposition 2.** *Let* $(X_n)_{n\in\mathbb{N}}$ *be an RCINAR($\infty$) with*

$$0 \leq \sum_{k=1}^{\infty} E[\alpha_k(Z_k)], \quad \sum_{k=1}^{\infty} |E[\alpha_k(Z_k)]| < 1,$$

*Then (2) has an almost surely unique first-order stationary solution* $(X_n)$ *where* $X_n \in \mathbb{N}_0$, $n \in \mathbb{Z}$, *and* $E[X_n] = \alpha_0/(1-K)$, *where* $K = \sum_{k=1}^{\infty} E[\alpha_k(Z_k)]$.

We can also give an AR($\infty$) representation of RCINAR($\infty$) process.

**Proposition 3** (AR($\infty$) representation of RCINAR($\infty$))**.** *Let* $(Z_k)_{k\in\mathbb{N}}$ *is an i.i.d. sequence of random variables,* $\alpha_0 > 0$, *and* $\alpha_k : \mathbb{R} \to \mathbb{R}$, $k = 1, 2, \ldots$ *are functions which satisfy the condition in Proposition 2, and let* $(X_n)$ *be the corresponding RCINAR($\infty$). Then*

$$u_n = X_n - \sum_{k=1}^{\infty} \alpha_k(Z_{n-k}) X_{n-k} - \alpha_0, \quad n \in \mathbb{Z}, \tag{3}$$

*defines a stationary sequence* $(u_n)$ *with* $E[u_n] = 0$ *and*

$$E[u_n u_m] = \begin{cases} 0 & n \neq m, \\ \frac{\alpha_0}{1-K} & n = m, \end{cases}$$

*where* $K = \sum_{k=1}^{\infty} E[\alpha_k(Z_k)]$.

**Remark 3.** Since it is difficult to simulate exact RCINAR($\infty$) process, we need to approximate the process by RCINAR($p$) with large $p$. Therefore, it is important to investigate the condition of first and second stationarity of RCINAR($p$). If $\alpha_k \equiv 0$ for $k > p$, then the random coefficient autoregressive process with infinite order (RCAR($\infty$)) (3) induced to RCAR($p$) process. In this case, let

$$Y_n = (X_n, X_{n-1}, \ldots, X_{n-p+1})^\top, \quad \xi_n = (u_n, 0, \ldots, 0)^\top,$$
$$c_0 = (a_0, 0, \ldots, 0)^\top, \quad A_n = \begin{pmatrix} a_1(Z_{n-1}) & a_2(Z_{n-2}) & \cdots & a_p(Z_{n-p}) \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & \vdots \\ \vdots & & & \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Then we can rewrite (3) in the form of RCAR(1) process

$$Y_n = c_0 + A_n Y_{n-1} + \xi_n. \tag{4}$$

Boshnakov (2011) investigates the necessary and sufficient condition for the existence of first- and second-order stationary (covariance stationary) solution of (4). In our case for the first-order stationarity, we require the condition

$$\mathrm{spr}(E[A_n]) < 1.$$

where $\mathrm{spr}(M)$ is the spectral radius, that is, the maximum modulus of eigenvalues of $M$ (see Kesten (1974) and Goldie and Maller (2000) for other related topics on stationarity of RCAR process). The assumptions in Proposition 2 implies that $\mathrm{spr}(E[A_n]) < 1$. For the second order stationarity, we require $\mathrm{spr}(E[A_n \otimes A_n]) < 1$ where $\otimes$ is the Hadamard product. This condition is also satisfied under assumptions in Proposition 2 and $\max_{1\le k\le p} E[\alpha_k(Z_k)^2] < 1$:

**Proposition 4.** *An RCINAR(p) process $(X_n)$ is covariance stationary if*

$$0 \le \sum_{k=1}^{p} E[\alpha_k(Z_k)], \quad \sum_{k=1}^{p} |E[\alpha_k(Z_k)]| < 1, \quad \max_{1\le k\le p} E[\alpha_k(Z_k)^2] < 1.$$

Therefore, the statement in Proposition 2 is consistent with the theoretical result on the stationarity of RCAR.

If the functions $\alpha_k$, $k = 0, 1, 2, \ldots$ are bounded and satisfy the conditions $|\alpha_k(x)| \le L_k$ for some $L_k > 0$ and $K_L = \sum_{k=1}^{\infty} L_k < 1$, then we have random coefficient moving-average (RCMA($\infty$)) representation and some properties autocovariance functions of RCINAR($\infty$). Proposition 5 plays an important role in the proof of Theorem 1.

**Proposition 5.** *Let $(X_n)$ be a RCINAR($\infty$) defined by (2) Suppose the conditions $|\alpha_k(x)| \le L_k$ for some $L_k > 0$, $k = 0, 1, 2, \ldots$, and $K_L = \sum_{k=1}^{\infty} L_k < 1$ holds. Then, for the autocovariance functions $R(j) = \mathrm{Cov}(X_n, X_{n+j})$, $j \in \mathbb{Z}$, we have*

$$\left| \sum_{j=0}^{\infty} R(j) \right| \le \frac{\alpha_0}{(1-K_L)^3} < \infty.$$

*In particular, we have $R(0) = \mathrm{Var}(X_n) \le \alpha_0/(1-K_L)^3$, $n \in \mathbb{Z}$.*

### 4.2 Discrete Approximation of SEPOT models

Next we give a limit theorem on the weak convergence of a marked point process constructed of the sum of a RCINAR($\infty$) to a marked Haweks process. This limit

theorem justify the approximation of a discretely sampled point process discussed in Section 2. As an extension of Proposition 1, we consider the following RCINAR($\infty$) process:

$$X_E^{\Delta_n}(k) = \epsilon_n(k) + \sum_{l=1}^{\infty} \left[ h\left(l\Delta_n\right) c\left(\widetilde{Z}_{k-l}\right) P\left( \frac{X_{(k-l)} - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \geq u \right) \right] \circ X_E^{\Delta_n}(k-l), \quad (5)$$

where $E = [u, \infty)$, $\widetilde{Z}_k = (Z_k - u)1_{\{Z_k \geq u\}}$, $Z_k \overset{i.i.d.}{\sim} GPD(\xi, 1)$, $c : \mathbb{R}_+ \to \mathbb{R}_+$ is a piecewise continuous function with $c(0) = 0$, $h : \mathbb{R} \to \mathbb{R}_+$ is also a piecewise continuous function with $h(t) = 0,\ t \leq 0,\ 0 \leq E[c(Z_1)] \int_{\mathbb{R}} h(t)dt < 1$, $X_k \overset{i.i.d.}{\sim} F \in D(G_\xi)$ and $\epsilon_n(k) \overset{i.i.d.}{\sim} Poi(\eta P(b_{[\Delta_n^{-1}]}(X_1 - a_{[\Delta_n^{-1}]}) \geq u))$ with $\eta > 0$. We assume $E[c(Z_1)^2] < \infty$ for technical reason and this assumption is satisfied if $0 < \xi < 1/2$ in Example 1 and satisfied in Example 2 if distribution $G$ have second moment.

We have the following relation between $\widetilde{X}_E^{\Delta_n}(k)$ and $\lambda(k\Delta_n | \mathcal{H}_{k\Delta_n}^N)$ by the definition of $X_E^{\Delta_n}$:

$$\begin{aligned}
E\left[ \frac{X_E^{\Delta_n}(k)}{\Delta_n} | \mathcal{F}_{(k-1)}^{\Delta_n} \right] &= \left( \eta + \sum_{m=1}^{\infty} h(k\Delta_n - m)c(\widetilde{Z}_{k-m}) X_E^{\Delta_n}(k-m) \right) \Delta_n^{-1} P\left( \frac{X_1 - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \geq u \right) \\
&= \left( \eta + \int_{-\infty}^{(k-1)\Delta_n} h(k\Delta_n - s)c(\widetilde{Z}_s) N_E^{\Delta_n}(ds) \right) \Delta_n^{-1} P\left( \frac{X_1 - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \geq u \right), \\
&\approx \left( \eta + \int_{-\infty}^{(k-1)\Delta_n} h((k-1)\Delta_n - s)c(\widetilde{Z}_s) N_g(ds) \right) (1 + \xi u)^{-1/\xi} \\
&= E\left[ \frac{N(\{(k-1)\Delta_n\} \times E)}{dt} | \mathcal{H}_{(k-1)\Delta}^N \right] = \lambda_E((k-1)\Delta_n | \mathcal{H}_{(k-1)\Delta_n}^N),
\end{aligned}$$

where $\widetilde{\mathcal{F}}_k^{\Delta_n} = \sigma((\widetilde{X}_{k-m}^{\Delta_n}, \widetilde{Z}_{k-m} : m \geq 0)$. Therefore, $X_E^{\Delta_n}$ and its conditional mean given $\mathcal{F}_{k-1}^{\Delta_n}$ can be interpret as an approximation of the increment $N(\{k\Delta_n\} \times E) - N(\{(k-1)\Delta_n\} \times E) = N(((k-1)\Delta_n, k\Delta_n] \times E)$ and the intensity function of $N_g$ respectively.

Next theorem gives a theoretical justification of this argument.

**Theorem 1.** *$(X_E^{\Delta_n}(k))$ be the RCINAR($\infty$) process defined by (5). Suppose following conditions are satisfied:*

$$c(x) \leq L, \quad L \int_{\mathbb{R}} h(t)dt < 1, \quad \textit{for some } L > 0,$$

*Consider the following marked point process*

$$\widetilde{N}^{\Delta_n}((a,b] \times E) = \widetilde{N}_E^{\Delta_n}((a,b]) = \sum_{k: k\Delta_n \in (a,b]} \widetilde{X}_E^{\Delta_n}(k), a < b, E = [u, \infty),$$

*where*

$$\widetilde{X}_E^{\Delta_n}(k) = \begin{cases} 1 & \text{if } X_E^{\Delta_n}(k) > 0, \\ 0 & \text{if } X_E^{\Delta_n}(k) = 0. \end{cases}$$

*Then we have* $\widetilde{N}_E^{\Delta_n} \xrightarrow{w} N$ *in* $M_p(\mathbb{R} \times E)$ *as* $\Delta_n \to 0$*. Here,* $N$ *is the marked Hawkes process with intensity*

$$\lambda_E(t|\mathcal{H}_t^N) = \left(\eta + \int_{-\infty}^{t} h(t-s)c(\widetilde{Z}_s)N(ds)\right)(1+\xi u)^{-1/\xi}, \tag{6}$$

*where* $\widetilde{Z}_s = (Z_s - u)^+$, $Z_s \overset{i.i.d.}{\sim} GPD(\xi, 1)$.

This result implies that RCINAR($\infty$) process (5) is the discrete time version of SEPOT models in Example 1 and 2 when the observation is high-frequency. The condition $c(x) \leq L$ for some $L > 0$ means that the contribution of mark to intensity function is bounded, and the condition $0 \leq L \int h(t)dt < 1$ enables us to investigate the relation between RCINAR($\infty$) and marked Hawkes process. This type of assumption is used in the literature of the stability of nonlinear Hawkes processes (see Brémaud and Massoulié (1996)). Marked point processes $\widetilde{N}_E^{\Delta_n}$ and $N_E^{\Delta_n}$ defined by $N^{\Delta_n}((a,b] \times E) = N_E^{\Delta_n}((a,b]) = \sum_{k:k\Delta_n \in (a,b]} X_E^{\Delta_n}(k)$ corresponds to $\widetilde{N}^{\Delta_n}$ and $N^{\Delta_n}$, which are defined in Section 2, respectively.

**Remark 4.** Theorem 1 is an extension of the results on the relation between INAR($\infty$) and Hawkes process in Kirchner (2016a) to the results on the relation between RCINAR($\infty$) and marked Hawkes process, and this theorem can also be interpreted as a generalization of Proposition 1. In fact, if we set $h(t) \equiv 0$ , $c(x) \equiv 1$ and $\eta = 1$ in the definition of $X_E^{\Delta_n}(k)$, then $X_E^{\Delta_n}(k) \sim Poi(\Delta_n)$ and this corresponds to Proposition 1.

**Remark 5.** In Theorem 1, a decay function $h$ is assumed to be non-negative. If decay function takes negative values, the limiting marked Hawkes process allows *self-damping*. In the literature of financial econometrics for example, we sometimes come across such a situation. This topic is considered and discussed in Section 5 and Section 6.

# 5 Simulation

In this section, we see the behavior of the weak convergence stated in Theorem 1. As an approximation of RCINAR($\infty$), we simulated RCINAR($p$) with large $p$. We set $p = 30$, $\widetilde{Z}_k \overset{i.i.d.}{\sim}$ GPD(0.2, 0.01), and for the kernel function $g$ in (1), we considered two functions

Case I : $g_1(t,z) = c(x)h_1(t) = [(1+x) \wedge 1.5]e^{-1.7t}$,

Case II: $g_2(t,z) = c(x)h_2(t) = [(1+x) \wedge 1.5]e^{-1.7t} \times \cos((1.5\pi t) \wedge 2\pi)$.

**Remark 6.** The bounded assumption in Cases I and II is not a restrictive condition in this case since $P(\widetilde{Z}_1 > 0.5) \approx 6.21 \times 10^{-6}$.

**Remark 7.** Case II corresponds to the case when a Hawkes process has *self-damping* property. This setting would be suggestive to see the asymptotic behavior of $N_E^{\Delta_n}$ when a decay function $h$ could take negative values. Self-damping is one of the recent important problem in financial econometrics. The problem is discussed in Section 6.

We also considered three cases for the mesh $\Delta_n = 1/4, 1/16, 1/32$ (we call these cases as A, B and C). In Figure 1, decay functions $h_1$ and $h_2$ are plotted. The function $h_1$ is always non-negative with exponential decay. On the other hand, the function $h_2$ takes negative values, and intuitively, this represents the situation that once the clustering of events observed, the following events less likely to occur for a while. Simulated values of $\{N_E^{\Delta_n}((0, k\Delta_n]) : k\Delta_n \in (0, 10]\}$, RCINAR($p$) $\{X_E^{\Delta_n}(k) : k\Delta_n \in (0, 10]\}$, and scaled conditional mean (conditional intensity) of $X_E^{\Delta_n}$ ($E = [u, \infty)$, $u = 0$), that is,

$$P_n^{-1} E\left[\frac{X_E^{\Delta_n}(k)}{\Delta_n} | \mathcal{F}_{(k-1)}^{\Delta_n}\right] = \eta + \sum_{m=1}^{\infty} h(k\Delta_n - m) c(\widetilde{Z}_{k-m}) X_E^{\Delta_n}(k-m), \quad k\Delta_n \in (0, 10],$$

where $P_n = \Delta_n^{-1} P\left(\frac{X_1 - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \geq u\right)$ are shown in Figures 2, 3 and 4. We first find that as the mesh $\Delta_n$ get small, $X_E^{\Delta_n}$ tends to take 0 or 1. This is quite natural because $X_E^{\Delta_n}$ is an approximation of the increment of the limiting simple point process in Theorem 1. We can see this in the center two figures in Figure 2, 3 and 4. Second, in Case II-A, B and C (these case are shown in the bottom right in Figures 2, 3 and 4), the scales conditional mean of $X_E^{\Delta_n}$ take negative value and values which are close to zero. This is because of the form of the decay function $h_2$: In Case II, $\int h_2^+(t)dt$ and $\int h_2^-(t)dt$ are close although $\int h_2^+(t)dt > \int h_2^-(t)dt$. Third, in Case I-C and II-C (the bottom two figures in Figure 4), the scales conditional mean of $X_E^{\Delta_n}$ is less smooth compared with Case I-B and II-B (the bottom two figures in Figure 3) in the interval that no events occur. This difference may come from the approximation of RCINAR($\infty$) by RCINAR($p$). This implies that we have to set more large $p$ as $\Delta_n$ goes to 0.

# 6 Discussion

In this section we first discuss about an extension of our results and then about an application of Theorem 1 to a statistical modeling of financial time series.

For the extension of *self-exciting* POT model to the multivariate case, we have to consider multivariate Hawkes process, which is called *mutually exciting* process, and a multivariate mark distribution (a joint jump size distribution). In the univariate SEPOT model, past events are usually considered to amplify the chance of occurrence of the same

type of events in some cases (the decay function $h$ is always nonnegative). However, in multivariate point process model, the event occurrence of a component could tend to reduce the event occurring probability of other components and in this case, decay function can be negative. This case is called *mutually-damping* and *self-damping* in the univariate case. Mutually-damping could happen in a high-frequency financial trading for example. Some trading activity reduces the possibility of future trading activity and has an adverse impact on its intensity. Boswijk, Laeven and Yang (2014) discuss the detection of self-excitation of events based on a general semimartingale theory. Eichler, Dahlhaus and Dueck (2016) and Kirchner (2016b) investigate nonparametric estimation of decay function of ordinary Hawkes process (Hawkes process with no marks), and in their real data analysis, some estimated decay functions take negative values. Therefore, self-(or mutually-)damping is a both theoretically and practically important problem. Moreover, the modeling of multivariate SEPOT model is related to the modeling of multivariate generalized Pareto distribution. These topics in the field of extreme value theory are presently under discussion. Falk and Guillou (2008) and Grothe, Korniichuk and Manner (2014) are important contributions in the theoretical and empirical standpoint of this topic.

For the statistical application of Theorem 1, Bayesian modeling of discrete-time SEPOT model may be possible. As noted in Kirchner (2016a) we can replace Poisson thinning operator in Definition 3 with Binomial thinning operator:

$$\alpha \circ X = \sum_{i=1}^{X} \xi_i, \quad \xi_i \overset{i.i.d.}{\sim} Bin(1, \alpha).$$

The *binomial* RCINAR would be convenient for MCMC methodology and may enable us to use non-i.i.d. jump size distribution. We refer to Neal and Rao (2007) for Bayesian estimation procedures of INAR($p$).

# 7 Conclusion

In this paper we introduced the general framework on the discrete observation of point process under the high-frequency observation. Grounded on this framework, we investigated the relation between RCINAR($\infty$) process and marked Hawkes process and gave a necessary and sufficient condition of the stationarity of RCINAR($\infty$) and its RCAR($\infty$) representation to build a bridge between the discrete-time series analysis and the analysis of continuous-time stochastic process. As applications of our results, we established the theoretical justification of self-exciting peaks over threshold models, which have been used in empirical financial time series analysis.

# A Proofs

We collect the proofs for Section 4. We use $\lesssim$, $\gtrsim$ to denote inequlities up to a multiplicative constant.

**Proof of Proposition 2.** We prove Proposition 2 in two steps.

**Step1:**(Existence of the stationary solution of (2)) First we construct a solution of (2). Let $\epsilon_i \overset{i.i.d.}{\sim} Poi(\alpha_0)$ and $Z_i \overset{i.i.d.}{\sim} F$. We define processes $(G_n^{(g,i,j)})$, $g \in \mathbb{N}_0$ recursively in the following procedure:

$$G_n^{(0,i,j)} \equiv 1_{\{n=0\}}, \quad n, i \in \mathbb{Z}, j \in \mathbb{N},$$
$$G_n^{(g,i,j)} \equiv \sum_{k=0}^{n} \alpha_k(Z_{n-k}) \circ G_{n-k}^{(g,i,j)}.$$

For $n < 0$, we set $G_n^{(g,i,j)} = 0$. Second, we define processes $(F_n^{i,j})$ as

$$F_n^{(i,j)} = \sum_{g=0}^{\infty} G_n^{g,i,j}, \quad n, i \in \mathbb{Z}, j \in \mathbb{N}.$$

Finally, we consider the process

$$\widetilde{X}_n \equiv \sum_{i=-\infty}^{n} \sum_{j=1}^{\epsilon_n} F_{n-i}^{(i,j)}, \quad n \in \mathbb{Z}.$$

It is possible to show that $(\widetilde{X}_n)$ solve (2) if we mimic the proof of Theorem 1 in Kirchner (2016a). One can also show that $E[\widetilde{X}_n] = \alpha_0/(1-K)$. In fact, since $E[\sum_{n=0}^{\infty} G_n^{(0,i,j)}] = 1$ and for $g > 0$,

$$\begin{aligned}
E\left[\sum_{n=0}^{\infty} G_n^{(g,i,j)}\right] &= \sum_{n=0}^{\infty} E\left[\sum_{k=1}^{n} \alpha_k(Z_{n-k}) \circ G_{n-k}^{(g,i,j)}\right] = \sum_{n=0}^{\infty} E\left[\sum_{k=1}^{n} \alpha_k(Z_{n-k}) G_{n-k}^{(g,i,j)}\right] \\
&= \sum_{n=0}^{\infty}\sum_{k=1}^{n} \widetilde{\alpha}_k E[G_{n-k}^{(g,i,j)}] = \sum_{k=0}^{\infty} \widetilde{\alpha}_k \sum_{n=k}^{\infty} E[G_{n-k}^{(g,i,j)}] = KE\left[\sum_{n=0}^{\infty} G_n^{(g,i,j)}\right] = \cdots \\
&= K^g.
\end{aligned}$$

Then we have

$$E[\widetilde{X}_n] = \alpha_0 \sum_{i-\infty}^{n} E[F_{n-i}^{i,j}] = \alpha_0 \sum_{g=0}^{\infty} E\left[\sum_{i=-\infty}^{n} G_{n-i}^{(g,i,j)}\right] = \alpha_0 \sum_{g=0}^{\infty} K^g = \frac{\alpha_0}{1-K}.$$

**Step2:**(Uniqueness of the solution of (2))

Let $(X_n)$ and $(Y_n)$ of (2) be two stationary solutions which are defined on the same probability space and with respect to the same immigration sequence $(\epsilon_n)$ and the same

offspring sequences with i.i.d.mark $(\xi_l^{n,k},\ Z_{n-k})$, $n \in \mathbb{Z}$, $k \in \mathbb{N}$. It follows from (2) that $E[|X_n|] = E[|Y_n|] = \alpha_0/(1-K) < \infty$. Then

$$\begin{aligned} E[|X_n - Y_n|] &= E\left|\sum_{k=1}^{\infty}(\alpha_k(Z_{n-k}) \circ X_{n-k} - \alpha_k(Z_{n-k}) \circ Y_{n-k})\right| \\ &\le \sum_{k=1}^{\infty} E\left|\alpha_k(Z_{n-k}) \circ X_{n-k} - \alpha_k(Z_{n-k}) \circ Y_{n-k}\right|. \end{aligned}$$

Then we also have

$$\begin{aligned} |\alpha_k(Z_{n-k}) \circ X_{n-k} - \alpha_k(Z_{n-k}) \circ Y_{n-k}| &= \left|\sum_{j=1}^{X_{n-k}} \xi_j^{n,k} - \sum_{j=1}^{Y_{n-k}} \xi_j^{n,k}\right| = \left|\sum_{j=X_{n-k}\wedge Y_{n-k}+1}^{X_{n-k}\vee Y_{n-k}} \xi_j^{n,k}\right| \\ &\overset{d}{=} \left|\sum_{j=1}^{|X_{n-k}-Y_{n-k}|} \xi_j^{n,k}\right| = \mathrm{I}_n. \end{aligned}$$

For $\mathrm{I}_n$ we have

$$\mathrm{I}_n = E[\alpha_k(Z_{n-k})]E|X_{n-k} - Y_{n-k}| = \widetilde{\alpha}_k E|X_n - Y_n|.$$

We used the stationarity of $(X_n)$ and $(Y_n)$ in the last inequality. Taking these together, we obtain that

$$E|X_n - Y_n| \le K E|X_n - Y_n|.$$

Since $K < 1$ by assumption and $E|X_n - Y_n| < \infty$, we obtain that $E|X_n - Y_n| = 0$ and therefore $X_n = Y_n$, $n \in \mathbb{Z}$ a.s. $\square$

**Proof of Proposition 3.** Let $\mathcal{F}_n = \sigma((X_{n-k}, Z_{n-k}) : k \le 0)$. Then we have

$$E[u_n] = E[E[u_n|\mathcal{F}_{n-1}]] = E\left[E[X_n|\mathcal{F}_{n-1}] - \alpha_0 - \sum_{k=1}^{\infty}\alpha_k(Z_{n-k})X_{n-k}\right] = 0.$$

If $n > m$, sence $E[u_n|\mathcal{F}_{n-1}] = 0$, we have

$$E[u_n u_m] = E[u_m E[u_n|\mathcal{F}_{n-1}]] = 0.$$

If $n = m$, we have

$$\begin{aligned} E[u_n u_n] &= \mathrm{Var}(u_n) = E[\mathrm{Var}(u_n|\mathcal{F}_{n-1})] + \mathrm{Var}(E[u_n|\mathcal{F}_{n-1}]) \\ &= E\left[\alpha_0 + \sum_{k=1}^{\infty}\alpha_k(Z_{n-k})X_{n-k}\right] = \alpha_0 + \frac{\alpha_0 K}{1-K} = \frac{\alpha_0}{1-K}. \end{aligned}$$

For the third equality, we used the facts $E[u_n|\mathcal{F}_{n-1}] = 0$, $\mathrm{Var}(u_n|\mathcal{F}_{n-1}) = \mathrm{Var}(X_n|\mathcal{F}_{n-1})$ and $X_n|\mathcal{F}_{n-1} \sim Poi(\alpha_0 + \sum_{k=1}^{\infty}\alpha_k(Z_{n-k})X_{n-k})$. This completes the proof. $\square$

**Proof of Proposition 4.** Let $\det(A)$ be the determinant of a matrix $A$ and $I_p$ be $p\times p$ unit matrix. The characteristic polynomial of the matrix $E[A_n]$ is given by

$$\det\left(\lambda I_p - E[A_n]\right) = \lambda^p - \alpha_1'\lambda^{p-1} - \cdots - \alpha_{p-1}'\lambda - \alpha_p',$$

where $\alpha_k' = E[\alpha_k(Z_k)]$, $k = 1,\dots,p$. Under assumptions of Proposition 4, one can see the maximum modulus of roots of this polynomial is less than 1 this implies $\mathrm{spr}(E[A_n]) < 1$. Now we show $\mathrm{spr}(E[A_n \otimes A_n]) < 1$. Since

$$\det(E[A_n \otimes A_n]) = \det(\widetilde{A}_n)\left(\det(E[A_n])\right)^{p-1},$$

where $\widetilde{A}_n = E[\alpha_p(Z_{n-p})A_n]$, it suffice to check $\mathrm{spr}(\widetilde{A}_n) < 1$. We can check this in the same way as the proof of $\mathrm{spr}(E[A_n]) < 1$. □

**Proof of Proposition 5.** Given the information of random coefficients $\mathcal{F}_Z = \sigma(Z_k : k \in \mathbb{Z})$, $X_E^{\Delta_n}$ behaves like AR($\infty$) under assumptions in Proposition 5. Let $\phi(z) = 1 - \sum_{k=1}^{\infty}\alpha_k(\cdot)z^k$ be a $\mathcal{F}_Z$-measurable function. Since $|\phi(z)| \geq 1 - \sum_{k\geq 0} L_k > 0$, we can define a random function $\psi(z) = \sum_{k=1}^{\infty}\beta_k z^k$ which satisfy the equation $\psi(z)\phi(z) = 1$. Comparing the coefficients of this equation, we have $\beta_0 = 1$, $\beta_k = \sum_{i=1}^{k}\alpha_i(\cdot)\beta_{k-i}$ and $\sum_{k=0}^{\infty}|\beta_k| = 1/\phi(1) < 1/(1-K_L)$ a.s. In this case, RCAR($\infty$) representation of $X_n$ given $\mathcal{F}_Z$ is invertible (see Theorem 3.1.2 in Brockwell and Davis (1991)) and have conditionally MA($\infty$) representation

$$X_n - \widetilde{\mu} = \sum_{k=0}^{\infty}\beta_k\widetilde{u}_{n-k}, \tag{7}$$

where $(\widetilde{u}_k)$ is the stationary sequence which satisfy the condition in Proposition 5. Let $\widetilde{R}(m) = \mathrm{Cov}(X_n, X_{n+m}|\mathcal{F}_Z)$, then from (7), we have

$$\left|\sum_{m=0}^{\infty}\widetilde{R}(m)\right| = \frac{\alpha_0}{1-\sum_{k=1}^{\infty}\alpha_k(\cdot)}\sum_{j=0}^{\infty}|\beta_j\beta_{j+m}| \leq \frac{\alpha_0}{1-K_L}\left(\sum_{k=0}^{\infty}|\beta_k|\right)^2 \leq \frac{\alpha_0}{(1-K_L)^3}.$$

Therefore, we have

$$\left|\sum_{m=0}^{\infty}R(j)\right| \leq E\left[\left|\sum_{m=0}^{\infty}\widetilde{R}(m)\right|\right] \leq \frac{\alpha_0}{(1-K_L)^3}.$$

□

Before we prove Theorem 1, we prepare some lemmas. Let $\mathcal{B}(\mathbb{R})$ be the Borel $\sigma$-field on $\mathbb{R}$, and $\mathcal{B}_b(\mathbb{R})(\subset \mathcal{B}(\mathbb{R}))$ be a family of bounded Borel sets.

**Lemma 1.** *For any $\Delta_n \in (0, \widetilde{\Delta})$, Let $N_E^{\Delta_n}$ be a marked point process defined as follows:*

$$N^{\Delta_n}((a,b] \times E) = N_E^{\Delta_n}((a,b]) = \sum_{k:k\Delta_n \in (a,b]} X_E^{\Delta_n}(k), a < b, E = [u, \infty), u \geq 0, \tag{8}$$

*where $(X_E^{\Delta_n}(k))_{k\in\mathbb{Z}}$ is the RCINAR($\infty$) process defined by (5). Then, for $A \in \mathcal{B}(\mathbb{R})$, we have that*

$$A \cap \{k\Delta_n : k \in \mathbb{Z}\} = \emptyset \Rightarrow N_E^{\Delta_n}(A) = 0 \quad a.s.$$

*For the expectation, we find that*

$$E[N_E^{\Delta_n}(\{k\Delta_n\})] = \frac{\eta_n}{1 - K_n} \leq \frac{\widetilde{\Delta}\eta(1 + \xi u)^{-1/\xi}}{1 - \widetilde{K}},$$

*where*

$$\eta_n = \eta P\left(\frac{X_1 - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \geq u\right), \quad K_n = E[c(\widetilde{Z}_1)]P\left(\frac{X_1 - a_{[\Delta_n^{-1}]}}{b_{[\Delta_n^{-1}]}} \geq u\right)\sum_{k=1}^{\infty} h(k\Delta_n),$$

*and where $E[c(\widetilde{Z}_1)] \int_0^T h(t)dt < \widetilde{K} < 1$.*

*Proof.* Lemma 1 follows from the definition of $N_E^{\Delta_n}$ and the stationarity of $X_E^{\Delta_n}$. □

**Lemma 2.** *The family of the probability measures $(P_n)$ on $(M_p(\mathbb{R} \times E), \mathcal{M}_p(\mathbb{R} \times E))$ corresponding to the point process $(N_E^{\Delta_n})_{\Delta_n \in (0,\widetilde{\Delta})}$ is uniformly tight.*

*Proof.* From Proposition 11.1.VI in Daley and Vere-Jones (2003), it is sufficient to show that for any compact interval $(a, b] \subset \mathbb{R}$ and for any $\epsilon > 0$, there exists $M > 0$ such that

$$\sup_{\Delta_n \in (0,\widetilde{\Delta})} P\left(N_E^{\Delta_n}((a,b]) > M\right) < \epsilon.$$

This results from Lemma 1 and Markov inequality:

$$P\left(N_E^{\Delta_n}((a,b]) > M\right) \leq \frac{E[N_E^{\Delta_n}((a,b])]}{M} \leq \frac{(b - a + 2\widetilde{\Delta})\eta(1 + \xi u)^{-1/\xi}}{M(1 - \widetilde{K})},$$

where $E[c(\widetilde{Z}_1)] \int_0^T h(t)dt < \widetilde{K} < 1$. By taking $M = (b - a + 2\widetilde{\Delta})\eta/(1 - \widetilde{K})\epsilon$, we have the desired result. □

**Lemma 3.** *Let $(N_E^{\Delta_n})_{\Delta_n \in (0,\widetilde{\Delta})}$, the family of point processes in Theorem 1 and $A \in \mathcal{B}_b(\mathbb{R})$. Then we have that the family of random variables $(N_E^{\Delta_n}(A))_{\Delta_n \in (0,\widetilde{\Delta})}$ is uniformly integrable.*

*Proof.* From the definition of $N_E^{\Delta_n}$ and Proposition 4, we have

$$\begin{aligned}
\mathrm{Var}(N_E^{\Delta_n}(A)) &= \mathrm{Var}\left(\sum_{k:k\Delta_n\in A} X_E^{\Delta_n}(k)\right) = \sum_{l:l\Delta_n\in A}\sum_{m:m\Delta_n\in A}\mathrm{Cov}(X_E^{\Delta_n}(l), X_E^{\Delta_n}(m)) \\
&\le \sum_{l:l\Delta_n\in A}\sum_{m=-\infty}^{\infty} R(l-m) = \sum_{k:k\Delta_n\in A}\sum_{m=-\infty}^{\infty} R(m) \\
&\lesssim \left(\frac{\sup A - \inf A}{\Delta_n}\right)\Delta_n < \infty.
\end{aligned}$$

This shows that $\mathrm{Var}(N_E^{\Delta_n}(A))$ is uniformly bounded in $\Delta_n \in (0,\widetilde{\Delta})$. Taking this and Lemma 2 together, for any $\epsilon > 0$, there exist $M = M_\epsilon > 0$ such that

$$\begin{aligned}
E[N_E^{\Delta_n}(A)1\{N_E^{\Delta_n} > M\}] &\lesssim \sup_{\Delta_n\in(0,\widetilde{\Delta})}\mathrm{Var}(N_E^{\Delta_n}(A)) \times \sup_{\Delta_n\in(0,\widetilde{\Delta})} P\left(N_E^{\Delta_n}(A) > M\right) \\
&\lesssim \epsilon.
\end{aligned}$$

This completes the proof. $\square$

Let $E$ is a complet separable metric space (c.s.m.s.) and $\mathcal{E}$ be a Borel $\sigma$-field on $E$. For any marked point process $N$ defined on $\mathbb{R}\times E$, we consider the following semiring $\mathcal{B}_a^N$:

$$\begin{aligned}
\mathcal{B}_a^N = \{&\{\omega\in\Omega : N((s_1,t_1]\times C_1)(\omega)\in D_1,\dots,N((s_k,t_k]\times C_k)(\omega)\in D_k\} : \\
&-\infty < s_i < t_i \le a, D_i\in\mathbb{N}_0, C_i\in\mathcal{E}, k\in\mathbb{N}\},
\end{aligned}$$

and let $\mathcal{H}_a^N$ be the $\sigma$-field generated by $\mathcal{B}_a^N$.

**Proof of Theorem 1.** We prove Theorem 1 in 3 steps.

**Step 1:**(Approximation of $\widetilde{N}_E^{\Delta_n}$ as $N_E^{\Delta_n}$) First we show $\widetilde{N}_E^{\Delta_n}((a,b]) - N_E^{\Delta_n}((a,b]) \xrightarrow{P} 0$ for all $a,b\in\mathbb{R}$ with $-\infty < a < b < \infty$. Consider the INAR($\infty$) processes:

$$X_{(1)}^{\Delta_n} = \epsilon_{n,(1)}(k) + \sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L)\circ X_{(1)}^{\Delta_n},$$

where $\epsilon_{n,(1)}(k) \overset{i.i.d.}{\sim} Poi(\Delta_n\eta)$, and define the following two auxiliary point processes:

$$N_{(1)}^{\Delta_n}((a,b]) = \sum_{k:k\Delta_n\in(a,b]} X_{(1)}^{\Delta_n}(k), \quad \widetilde{N}_{(1)}^{\Delta_n}((a,b]) = \sum_{k:k\Delta_n\in(a,b]} \widetilde{X}_{(1)}^{\Delta_n}(k), \quad a<b,$$

where

$$\widetilde{X}_{(1)}^{\Delta_n}(k) = \begin{cases} 1 & \text{if } X_{(1)}^{\Delta_n}(k) > 0 \\ 0 & \text{if } X_{(1)}^{\Delta_n}(k) = 0. \end{cases}$$

From the definitions of $N_E^{\Delta_n}$, $\widetilde{N}_E^{\Delta_n}$, $N_{(1)}^{\Delta_n}$ and $\widetilde{N}_{(1)}^{\Delta_n}$, we have

$$
\begin{aligned}
P\left(\left|N_E^{\Delta_n}((a,b]) - \widetilde{N}_E^{\Delta_n}((a,b])\right| > \epsilon\right) &\leq P\left(\left|N_{(1)}^{\Delta_n}((a,b]) - \widetilde{N}_{(1)}^{\Delta_n}((a,b])\right| > \epsilon\right) \\
&\leq P\left(\bigcup_{k:k\Delta_n \in (a,b]} \{X_{(1)}^{\Delta_n}(k) \geq 2\}\right) \\
&\leq \sum_{k:k\Delta_n \in (a,b]} P\left(X_{(1)}^{\Delta_n}(k) \geq 2\right).
\end{aligned}
$$

Moreover, let $\mathcal{F}_k^{\Delta_n} = \sigma(X_{(1)}^{\Delta_n}(l), l \leq k)$. Then we have

$$
\begin{aligned}
P\left(X_{(1)}^{\Delta_n}(k) \geq 2|\mathcal{F}_{k-1}^{\Delta_n}\right) &\leq P\left(\epsilon_{n,(1)}(k) \geq 2, \sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L) \circ X_{(1)}^{\Delta_n}(k-l) \geq 0|\mathcal{F}_{k-1}^{\Delta_n}\right) \\
&\quad + P\left(\epsilon_{n,(1)}(k) \geq 1, \sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L) \circ X_{(1)}^{\Delta_n}(k-l) \geq 1|\mathcal{F}_{k-1}^{\Delta_n}\right) \\
&\quad + P\left(\epsilon_{n,(1)}(k) \geq 0, \sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L) \circ X_{(1)}^{\Delta_n}(k-l) \geq 2|\mathcal{F}_{k-1}^{\Delta_n}\right) \\
&\leq P(\epsilon_{n,(1)}(k) \geq 2) + P(\epsilon_{n,(1)}(k) \geq 1)E[X_{(1)}^{\Delta_n}(k)] \\
&\quad + P\left(\sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L) \circ X_{(1)}^{\Delta_n}(k-l) \geq 2|\mathcal{F}_{k-1}^{\Delta_n}\right) \\
&\lesssim \Delta_n^2 + P\left(\sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L) \circ X_{(1)}^{\Delta_n}(k-l) \geq 2|\mathcal{F}_{k-1}^{\Delta_n}\right),
\end{aligned}
$$

where we used $P(\epsilon_{n,(1)}(k) \geq 2) \lesssim \Delta_n^2$, $P(\epsilon_{n,(1)}(k) \geq 1) \lesssim \Delta_n$ and $E[X_{(1)}^{\Delta_n}(k)] \lesssim \Delta_n$. Since

$$
\sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L) \circ X_{(1)}^{\Delta_n}(k-l)|\mathcal{F}_{k-1}^{\Delta_n} \sim Poi\left(\lambda_{n,k}\right),
$$

where $\lambda_{n,k} = \sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L)\, X_{(1)}^{\Delta_n}(k-l)$, we also have

$$
P\left(\sum_{l=1}^{\infty}(\Delta_n h(l\Delta_n)L) \circ X_{(1)}^{\Delta_n}(k-l) \geq 2|\mathcal{F}_{k-1}^{\Delta_n}\right) \lesssim \lambda_{n,k}^2,
$$

and

$$
\begin{aligned}
E[\lambda_{n,k}^2] &= E\left[\sum_{l=1}^{\infty} \Delta_n^2 h^2(l\Delta_n) L^2 (X_{(1)}^{\Delta_n}(k-l))^2\right] \\
&\quad + E\left[\sum_{l\neq m} \Delta_n^2 h(l\Delta_n) h(m\Delta_n) L^2 X_{(1)}^{\Delta_n}(k-l) X_{(1)}^{\Delta_n}(k-m)\right] \\
&= \sum_{l=1}^{\infty} \Delta_n^2 h^2(l\Delta_n) L^2 \left[\mathrm{Var}(X_{(1)}^{\Delta_n}(k-l)) + (E[X_{(1)}^{\Delta_n}(k-l)])^2\right] \\
&\quad + \sum_{l\neq m} \Delta_n^2 h(l\Delta_n) h(m\Delta_n) L^2 \left[\mathrm{Cov}(X_{(1)}^{\Delta_n}(k-l), X_{(1)}^{\Delta_n}(k-m)) + E[X_{(1)}^{\Delta_n}(k-l)] E[X_{(1)}^{\Delta_n}(k-m)]\right] \\
&\lesssim \Delta_n^2 \left(\Delta_n L \sum_{l=1}^{\infty} |h(l\Delta_n)|\right) + (\sup|h|)\Delta_n \left(\Delta_n L \sum_{l=1}^{\infty} |h(l\Delta_n)|\right)\left(\sum_{m=0}^{\infty} \mathrm{Cov}(X_{(1)}^{\Delta_n}(-l), X_{(1)}^{\Delta_n}(-m))\right) \\
&\lesssim \Delta_n^2.
\end{aligned}
$$

Here, we used $E[X_{(1)}^{\Delta_n}(k)] \lesssim \Delta_n$, $k \in \mathbb{Z}$, $\mathrm{Var}(X_{(1)}^{\Delta_n}(k)) \lesssim \Delta_n$, $k \in \mathbb{Z}$, and $\sum_{m=0}^{\infty} \mathrm{Cov}(X_{(1)}^{\Delta_n}(0), X_{(1)}^{\Delta_n}(m)) \lesssim \Delta_n$ which are obtained from Proposition 5. Therefore, we have for sufficiently small $\Delta_n$,

$$
\sum_{k:k\Delta_n \in (a,b]} P\left(N_{(1)}^{\Delta_n}(((k-1)\Delta_n, k\Delta_n]) \geq 2\right) \lesssim \Delta_n = o(1).
$$

Hence $P\left(\left|N_E^{\Delta_n}((a,b]) - \widetilde{N}_E^{\Delta_n}((a,b])\right| > \epsilon\right) \to 0, \Delta_n \to 0, \forall \epsilon > 0$. This establishes $\widetilde{N}_E^{\Delta_n}((a,b]) - N_E^{\Delta_n}((a,b]) \xrightarrow{P} 0$. Then, it suffice to show $N_E^{\Delta_n}((a,b]) \xrightarrow{w} N((a,b] \times E)$.

Now we show $N_E^{\Delta_n} \xrightarrow{w} N$ where $N_E^{\Delta_n}$ is the point process defined by (8). The marked Hawkes process $N$ solve the equation

$$
\begin{aligned}
&E[1_{A^*} N^*((a,b] \times E)] \\
&\quad = E\left[1_{A^*} \int_a^b \left(\eta + \int_{-\infty}^t \int_E h(t-s) c(z) N^*(ds \times dz)\right) dt\right], a < b, A^* \in \mathcal{H}_a^{N^*}. \qquad (9)
\end{aligned}
$$

It is known that if $E[c(Z_1)] \int_0^\infty |h(t)| dt \leq L \int_0^\infty |h(t)| dt < 1$, there exist a unique stationary solution $N$ of (9) (see Brémaud, Nappo and Torrisi (2002) for the sufficient condition of stationarity and Kerstan (1964) for the existence and uniqueness of stationary solution, which is a marked Hawkes process defined by the intensity function (6)). Let $P_n = \Delta_n^{-1} P\left(\frac{X_1 - a_{[\Delta_n]^{-1}}}{b_{[\Delta_n]^{-1}}} \geq u\right)$. In the following proof, we omit the index $E$ of $N_E^{\Delta_n}$ and $X_E^{\Delta_n}$ for convenience. For $N^{\Delta_n}$, we have

$$
\begin{aligned}
&E[1_{A_n} N^{\Delta_n}((a,b] \times E)] \\
&\quad = \Delta_n \sum_{k:k\Delta_n \in (a,b]} E\left[1_{A_n} P_n \left(\eta + \int_{-\infty}^{k\Delta_n} \int_E h(k\Delta_n - s) c(z) N^{\Delta_n}(ds \times dz)\right)\right], \quad A_n \in \mathcal{B}_a^{N^{\Delta_n}}
\end{aligned}
$$

Since random variables in the expectation can be written as $\Phi(N^{\Delta_n})$ for some measurable map $\Phi$ and it is possible to show $P(N^* \in D_\Phi) = 0$ for the marked point process $N^*$ which solve (9) and the set of discontinuous points of $\Phi$, $D_\Phi$ (see the proof of Theorem 2 in Kirchner (2016a)). Therefore, from Lemma 2 and continuous mapping theorem, we have $1_{A_n} N^{\Delta_n}((a,b] \times E) \xrightarrow{w} 1_{A^*} N^*((a,b] \times E)$ and

$$\lim_{n\to\infty} E[1_{A_n} N^{\Delta_n}((a,b] \times E)] = E[1_{A^*} N^*((a,b] \times E)].$$

Therefore, for the weak convergence of hole sequence $(N^{\Delta_n})_{\Delta_n \in (0,\widetilde{\Delta})}$ to $N^*$, it is sufficient to show

$$\begin{aligned}\lim_{n\to\infty} \Delta_n \sum_{k:k\Delta_n \in (a,b]} & E\left[1_{A_n}\left(\eta + \int_{-\infty}^{k\Delta_n}\int_E h(k\Delta_n - s)c(z)N^{\Delta_n}(ds \times dz)\right)\right] \\ &= \int_a^b E\left[1_{A^*}\left(\eta + \int_{-\infty}^{t}\int_E h(t-s)c(z)N^*(ds \times dz)\right)\right] dt \qquad (10)\end{aligned}$$

and this establishes our desired result since the law of point process is uniquely determined by its intensity process. We show (10) in several steps.

**Step 2:**(Uniform integrability of $\left(\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds \times dz)\right)_{\Delta_n \in (0,\widetilde{\Delta})}$)

Let $M > 0$ with $-M < a$ be a constant and $\mathcal{K}$ be a compact subset of $E$. If $\mathrm{Var}\left(\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds \times dz)\right)$ is uniformly bounded in $\Delta_n$, we can show that $\left(\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds \times dz)\right)_{\Delta_n \in (0,\widetilde{\Delta})}$ is uniformly tight. Then by continuous mapping theorem, for $-M < a$, we have

$$1_{A_n}\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds \times dz) \xrightarrow{w} 1_{A^*}\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^*(ds \times dz).$$

Taking this with the uniform integrability of $\left(1_{A_n}\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds \times dz)\right)_{\Delta_n \in (0,\widetilde{\Delta})}$, we obtain

$$\lim_{n\to\infty} E\left[1_{A_n}\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds \times dz)\right] = E\left[1_{A^*}\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^*(ds \times dz)\right].$$

Now we show uniform boundedness of $\mathrm{Var}\left(\int_{-M}^t \int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds \times dz)\right)$.

$$
\begin{aligned}
&\mathrm{Var}\left(\int_{-M}^{t}\int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds\times dz)\right)\\
&\quad\le \sum_{l=1}^{[M/\Delta_n]+1}\sum_{m=1}^{[M/\Delta_n]+1} |h(l\Delta_n)||h(m\Delta_n)|\,\mathrm{Cov}(c(\widetilde{Z}_l)X^{\Delta_n}(l), c(\widetilde{Z}_m)X^{\Delta_n}(m))\\
&\quad\le \sum_{l=1}^{[M/\Delta_n]+1}\sum_{m=1}^{[M/\Delta_n]+1} |h(l\Delta_n)||h(m\Delta_n)|\left[\mu_{X^{\Delta_n}}^2\,\mathrm{Cov}(c(\widetilde{Z}_l), c(\widetilde{Z}_m)) + \mu_z^2\,\mathrm{Cov}(X^{\Delta_n}(l), X^{\Delta_n}(m))\right]\\
&\quad= (\sup|h|)^2\mu_{X^{\Delta_n}}^2 \sum_{l=1}^{[M/\Delta_n]+1}\sum_{m=1}^{[M/\Delta_n]+1}\mathrm{Cov}(c(\widetilde{Z}_l), c(\widetilde{Z}_m))\\
&\quad+ (\sup|h|)^2\mu_z^2 \sum_{l=1}^{[M/\Delta_n]+1}\sum_{m=1}^{[M/\Delta_n]+1}\mathrm{Cov}(X^{\Delta_n}(l), X^{\Delta_n}(m))\\
&\quad\lesssim \mu_{X^{\Delta_n}}^2\Delta_n^2 + \sum_{m=0}^{\infty} R(m) \lesssim \mu_{X^{\Delta_n}}^2\Delta_n^2 + \Delta_n < \infty.
\end{aligned}
$$

where $\mu_z = E[c(\widetilde{Z}_1)]$, $\mu_{X^{\Delta_n}} = \Delta_n\eta/(1-K)$ and $R(m) = \mathrm{Cov}(X^{\Delta_n}(0), X^{\Delta_n}(m))$. For the last inequality, we used Proposition 5. Then, we have

$$
\sup_{\Delta_n\in(0,\widetilde{\Delta})} \mathrm{Var}\left(\int_{-M}^{t}\int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds\times dz)\right) < \infty
$$

uniformly in $\Delta_n \in (0,\widetilde{\Delta})$. Therefore, random variables $1_{A_n}\int_{-M}^{t}\int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds\times dz)$, $n\in\mathbb{N}$ are uniformly integrable. Together with weak convergence, we established that for $M$ with $-M<a$,

$$
\lim_{n\to\infty} E\left[1_{A_n}\int_{-M}^{t}\int_{\mathcal{K}} h(t-s)c(z)N^{\Delta_n}(ds\times dz)\right] = E\left[1_{A^*}\int_{-M}^{t}\int_{\mathcal{K}} h(t-s)c(z)N^{*}(ds\times dz)\right].
$$

**Step3:**(Evaluation of reminder terms) Let $P_n = \Delta_n^{-1}P(b_n^{-1}(X_1-a_n)\ge u)$. We decompose

$$
\begin{aligned}
&\Bigg|\sum_{k:k\Delta_n\in(a,b]} P_n\Delta_n E\left[1_{A_n}\int_{-\infty}^{k\Delta_n}\int_E h(k\Delta_n-s)N^{\Delta_n}(ds\times dz)\right]\\
&\qquad -(1+\xi u)^{-1/\xi}\int_a^b E\left[1_{A^*}\int_{-\infty}^{t}\int_E h(t-s)c(z)N^{*}(ds\times dz)\right]dt\Bigg|
\end{aligned}
$$

into the following five terms:

$$
\begin{aligned}
\mathrm{I}_n &= \Bigg| \sum_{k:k\Delta_n \in (a,b]} P_n \Delta_n E\left[1_{A_n} \int_{-M}^{k\Delta_n} \int_{\mathcal{K}} h(k\Delta_n - s) c(z) N^{\Delta_n}(ds \times dz)\right] \\
&\qquad - (1+\xi u)^{-1/\xi} \int_a^b E\left[1_{A^*} \int_{-M}^{t} \int_{\mathcal{K}} h(t-s) c(z) N^*(ds \times dz)\right] dt \Bigg|, \\
\mathrm{II}_n &= \sum_{k:k\Delta_n \in (a,b]} P_n \Delta_n E\left[1_{A_n} \int_{-M}^{k\Delta_n} \int_{\mathcal{K}^c} h(k\Delta_n - s) c(z) N^{\Delta_n}(ds \times dz)\right], \\
\mathrm{III}_n &= (1+\xi u)^{-1/\xi} \int_a^b E\left[1_{A^*} \int_{-M}^{t} \int_{\mathcal{K}^c} h(t-s) c(z) N^*(ds \times dz)\right] dt, \\
\mathrm{IV}_n &= \sum_{k:k\Delta_n \in (a,b]} P_n \Delta_n E\left[1_{A_n} \int_{-\infty}^{-M} \int_{E} h(k\Delta_n - s) c(z) N^{\Delta_n}(ds \times dz)\right], \\
\mathrm{V}_n &= \int_a^b E\left[1_{A^*} \int_{-\infty}^{-M} \int_{E} h(t-s) c(z) N^*(ds \times dz)\right] dt.
\end{aligned}
$$

From a similar argument of the proof of Theorem 2 in Kirchner (2016a), it is possible to show that for any $\epsilon > 0$, there exist $M = M_\epsilon$ such that

$$
\mathrm{I}_n < \frac{\epsilon}{5}, \quad |\mathrm{IV}_n| < \frac{\epsilon}{5}, \quad |\mathrm{V}_n| < \frac{\epsilon}{5}. \tag{11}
$$

For $\mathrm{II}_n$, we have

$$
\begin{aligned}
\left| E\left[1_{A_n} \int_{-M}^{k\Delta_n} \int_{\mathcal{K}^c} h(k\Delta_n - s) c(z) N^{\Delta_n}(ds \times dz)\right] \right| &\le E\left[\int_{-M}^{b} \int_{\mathcal{K}^c} |h(t-s)| c(z) N^{\Delta_n}(ds \times dz)\right] \\
&\le (\sup|h|) E[c(\widetilde{Z}_1) : \mathcal{K}^c] \sum_{k:k\Delta_n \in (-M,b]} E[X_E^{\Delta_n}(k)] \\
&\le (\sup|h|) \mu_Z^2 P(Z_1 \in \widetilde{\mathcal{K}}^c) \frac{([M+b]+1)\eta}{1-K},
\end{aligned}
$$

where $E[c(\widetilde{Z}_1) : \mathcal{K}^c]$ means the expectation of $c(\widetilde{Z}_1)$ over $\mathcal{K}^c$. Since $Z_1$ have tight law, for any $\widetilde{\epsilon} > 0$ and $M > 0$, we can take a compact set $\widetilde{\mathcal{K}} = \widetilde{\mathcal{K}}_{\widetilde{\epsilon},M} \subset [0,\infty)$ such that

$$
P(Z_1 \in \widetilde{\mathcal{K}}^c) < \frac{\widetilde{\epsilon}(1-\widetilde{K})}{5\eta(\sup|h|)\mu_Z^2([M+b]+1)},
$$

where $E[c(Z_1)] P(b^{-1}_{[\Delta_n]^{-1}}(X_1 - a_{[\Delta_n]^{-1}}) \ge u) \sum_{k=1}^{\infty} |h(k\Delta_n)| < \widetilde{K} < 1$. Then we have

$$
\left| E\left[1_{A_n} \int_{-M}^{k\Delta_n} \int_{\mathcal{K}^c} h(k\Delta_n - s) c(z) N^{\Delta_n}(ds \times dz)\right] \right| < \frac{\epsilon}{5}. \tag{12}
$$

For III$_n$, since we have $E[\int_{\mathcal{K}} c(z)N^*(dt \times dz)/dt] \le \frac{E[c(\widetilde{Z}_1):\mathcal{K}]\eta}{(1-\widetilde{K})}$, for any compact set $\mathcal{K} \in \sigma(\mathcal{Z})$($\sigma$-field associated to $\mathcal{Z}$), we have

$$\begin{aligned}\left|E\left[1_{A^*}\int_{-M}^{t}\int_{\mathcal{K}^c} h(k\Delta_n - s)c(z)N^*(ds\times dz)\right]\right| &\le E\left[\int_{-M}^{t}\int_{\mathcal{K}^c} |h(t-s)|c(z)N^*(ds\times dz)\right]\\ &\le \frac{E[c(\widetilde{Z}_1):\mathcal{K}^c]\eta}{1-\widetilde{K}}\int_0^{b+M}|h(t)|dt\\ &\le \frac{E[c(\widetilde{Z}_1)^2]P(Z_1\in\widetilde{\mathcal{K}}^c)\eta}{1-\widetilde{K}}\int_0^{\infty}|h(t)|dt\end{aligned}$$

Therefore we established (10). From the same technique used in the evaluation of II$_n$, we also have

$$\left|E\left[1_{A^*}\int_{-M}^{t}\int_{\mathcal{K}^c} h(t-s)c(z)N^*(ds\times dz)\right]\right| < \frac{\epsilon}{5}. \tag{13}$$

Taking (11), (12) and (13) together, we have

$$\begin{aligned}&\lim_{n\to\infty}\Delta_n\sum_{k:k\Delta_n\in(a,b]}E\left[1_{A_n}\left(\eta+\int_{-\infty}^{k\Delta_n}\int_E h(k\Delta_n - s)c(z)N^{\Delta_n}(ds\times dz)\right)\right]\\ &=\int_a^b E\left[1_{A^*}\left(\eta+\int_{-\infty}^{t}\int_E h(t-s)c(z)N^*(ds\times dz)\right)\right]dt.\end{aligned}$$

Then we established the desired result. $\square$

# B Figures and tables


Figure 1: Left: $h_1$, right: $h_2$

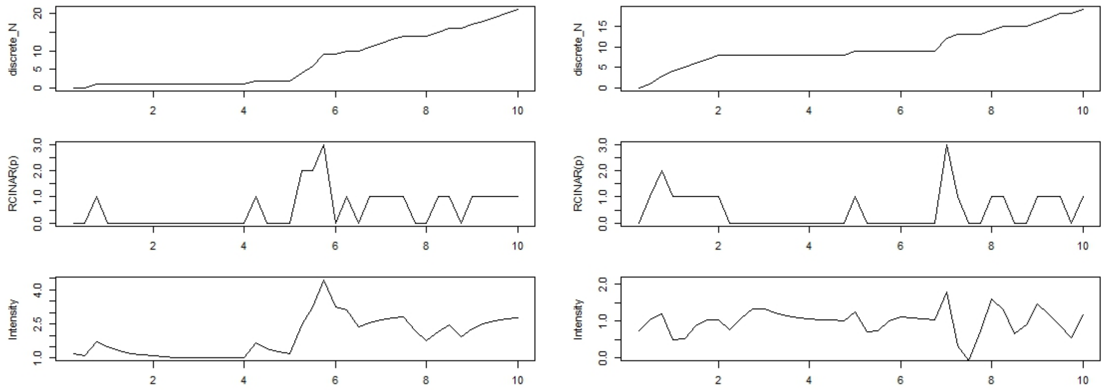


Figure 2: Left: Case I-A, right: Case II-A. Top: $N_E^{\Delta_n}$, center: RCINAR$(p)$, bottom: conditional mean of $X_E^{\Delta_n}$.

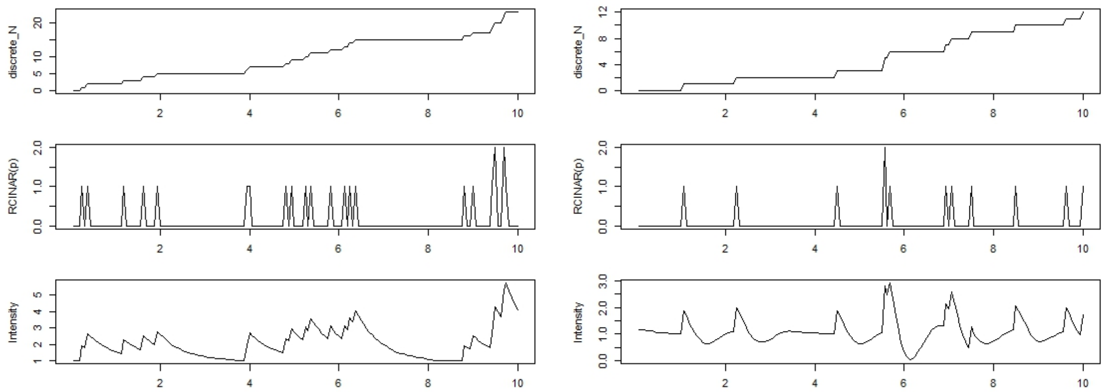


Figure 3: Left: Case I-B, right: Case II-B. Top: $N_E^{\Delta_n}$, center: RCINAR$(p)$, bottom: conditional mean of $X_E^{\Delta_n}$.

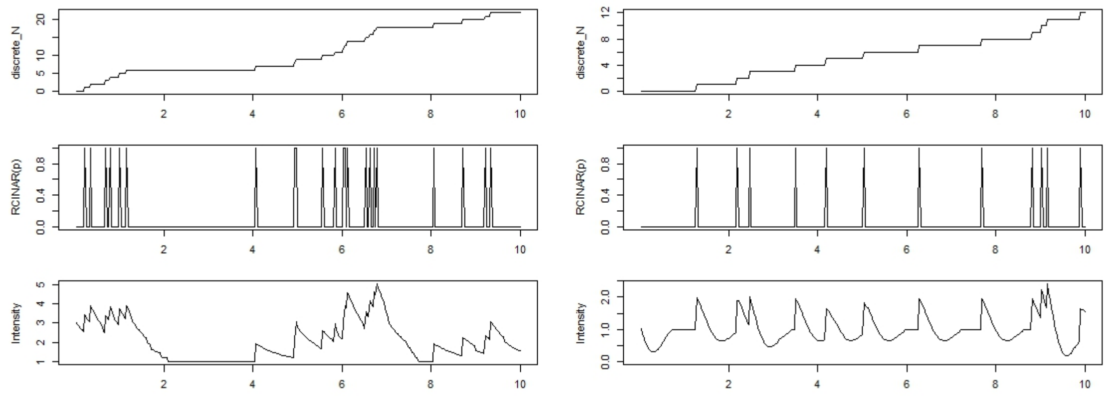


Figure 4: Left: Case I-C, right: Case II-C. Top: $N_E^{\Delta_n}$, center: RCINAR($p$), bottom: conditional mean of $X_E^{\Delta_n}$.

# References


Aït-Sahalia, Y. and Jacod, J. (2014), High-frequency financial econometrics. *Princeton Univ. Press.*

Al-Osh, M. and Alzaid, A.A. (1987), First-order integer-valued autoregressive (INAR(1)) process, *J. Time Ser. Anal.* 8, 261-275.

Alzaid, A.A. and Al-Osh, M. (1990), An integer-valued pth-order autoregressive structure (INAR(p)) process, *J. Appl. Probab.* 27-2, 314-324.

Balkema, A.A. and de Haan, L. (1974), Residual life time at great age. *Ann. Probab.* 2, 792-804.

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley and Sons.

Boshnakov, G.N. (2011), On first and second stationarity of random coefficient models, *Linear Algebra and its Appl.* 434, 415-423.

Boswijk, H.P., Laeven, R.J.A. and Yang, X. (2014), Testing for self-excitation in jumps. working paper.

Brémaud, P. and Massoulié, L. (1996), Stability of nonlinear Hawkes processes. *Ann. Probab.* 24-3, 1563-1588.

Brémaud, P., Nappo, G. and Torrisi, G.L. (2002), Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Probab.* 39, 123-136.

Brockwell, P.J. and Davis, R.A. (1991), *Time Series: Theory and Methods*, 2nd Edition, Springer.

Chavez-Demoulin, V., Davison, A.C. and McNeil, A.J. (2005), A point process approach to Value-at-Risk estimation, *Quant. Fin.* 5-2, 227-234.

Chavez-Demoulin, V., Embrechts, P. and Sardy, S. (2014), Extreme-quantile tracking for financial time series. *J. Econometrics.* 181, 44-52.

Chavez-Demoulin, V. and McGill, J.A. (2012), High-frequency financial data modeling using Hawkes processes. *J. Banking Finance.* 36, 3415-3426.

Daley, D.J. and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes*, Volume I, 2nd, Edition, Springer.

de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer.

Du, J. G. and Li, Y. (1991), The integer-valued autoregressive (INAR(p)) model, *J. Time Ser. Anal.*, 12-2, 129-142.

Eichler, M., Dahlhaus, R. and Dueck, J. (2016), Graphical modeling for multivariate Hawkes processes with nonparametric link functions. To appear in *J. Time Ser. Anal.*

Falk, M. and Guillou, A. (2008), Peaks-over-threshold stability of multivariate generalized Pareto distributions. *J. Multivar. Anal.* 99, 715-734.

Goldie, C.M. and Maller, R. (2000), Stability of perpetuities, *Ann. Probab.* 28, 1195-1218.

Grothe, O., Korniichuk, V. and Manner, H. (2014), Modeling multivariate extreme events using self-exciting point processes. *J. Econometrics* 182, 269-289.

Hawkes, A.G. (1971), Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58-1, 83-90.

Kerstan, J. (1964), Teliprozesse Possonscher Prozesse. In *Trans. 3rd Prague Conf. Inf. Theory, Statist. Decision Functions, Random Process.* (Liblice, 1962), Czechoslovak Academy of Science, Prague, 377-403.

Kesten, H. (1974), Renewal theory for functional of a Markov chain with general state space. *Ann. Probab.* 2, 355-386.

Kirchner, M. (2016a), Hawkes and INAR($\infty$) processes, *Stoch. Proc. Appl.*, 126, 2494-2525.

Kirchner, M. (2016b), An estimation procedure for the Hawkes process. *Quant. Fin.* doi: 10.1080/14697688.2016.1211312.

Kunitomo, N. Ehara, A. and Kurisu, D. (2016), Causality analysis of financial markets by using the multivariate Hawkes type models. Discussion Paper CIRJE-J-278 (in Japanese).

Liniger, T.J. (2009), Multivariate Hawkes Processes, Ph.D. thesis.

McCabe, B.P.M., Martin, G.M. and Harris, D.G. (2011), Efficient probabilistic forecasts of counts. *J. R. Stat. Soc. Ser. B.* 73, 253-272.

McKenzie, E. (1985), Some simple models for discrete variate time series, *Water Resources Bulletin*, 21, 645-650.

Mckenzie, E. (2003), *Discrete variate time series.* In: Shanbhag, D.N. and Rao, C.R. (eds), Handbook of Statistics, Elsevier Science, pp. 573-606.

McNeil, A.J., Frey, R. and Embrechts, P. (2005), *Quantitative Risk Management: Concepts, Techniques And Tools*, Princeton University Press.

Neal, P. and Rao, S.T. (2007), MCMC for integer-valued ARMA processes. *J. Time Ser. Anal.* 28-1, 92-110.

Nicholls, D.F. and Quinn, B.G. (1982), *Random-Coefficient Autoregressive Models: An Introduction.* Lecture Notes in Statist. 11. Springer, New York.

Pickands III, J. (1975), Statistical inference using extreme order statistics. *Ann. Statist.* 3, 119-131.

Resnick, S.I. (1987), *Extreme Values, Regular Variation, and Point Processes.* Springer.

Resnick, S.I. (2007), *Heavy Tail Phenomena*, Springer.

Zheng, H., Basawa, I.V. and Datta, S. (2007), First-order random coefficient integer-valued autoregressive processes. *J. Statist. Plann. Inference.* 173, 212-229.

Zhang, H., and Wang, D. and Zhu, F. (2011a), Empirical likelihood inference for random coefficient INAR($p$) process. *J. Time Ser. Anal.* 32, 195-203.

Zhang, H., and Wang, D. and Zhu, F. (2011b), The empirical likelihood for first-order random coefficient integer-valued autoregressive processes. *Comm. Statist. Theory Methods.* 40, 492-509.